

## TITLE: DOCUMENT DESCRIPTOR EXTRACTION METHOD

## ABSTRACT OF THE DISCLOSURE

A document descriptor extraction method and system which creates a document  
5 descriptor by generalizing input sequences within a document; factoring the input sequences  
and generalized input sequences; and selecting a document descriptor from the input  
sequences, generalized sequences, and factored sequences, preferably using minimum  
descriptor length (MDL) principles. Novel algorithms are employed to perform the  
generalizing, factoring, and selecting.

M:\SWeed\Lucent Technologies\23,760\Patent Office\application